PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
H04L 12/18, 29/14

A1

(11) International Publication Number: **WO 99/48246**

(43) International Publication Date: 23 September 1999 (23.09.99)

(21) International Application Number: PCT/US99/05731

(22) International Filing Date: 16 March 1999 (16.03.99)

(30) Priority Data:
09/039,375 16 March 1998 (16.03.98) US

(71) Applicant: REUTERS AMERICA INCORPORATED [US/US]; 88 Parkway Drive South, Hauppauge, NY 11788 (US).

(72) Inventor: STIRPE, Paul, A.; 3 Kristina Way, Shoreham, NY 11786 (US).

(74) Agents: GLEMBOCKI, Christopher, R. et al.; Banner & Witcoff, Ltd., 11th floor, 1001 G Street, N.W., Washington, DC 20001-4597 (US).

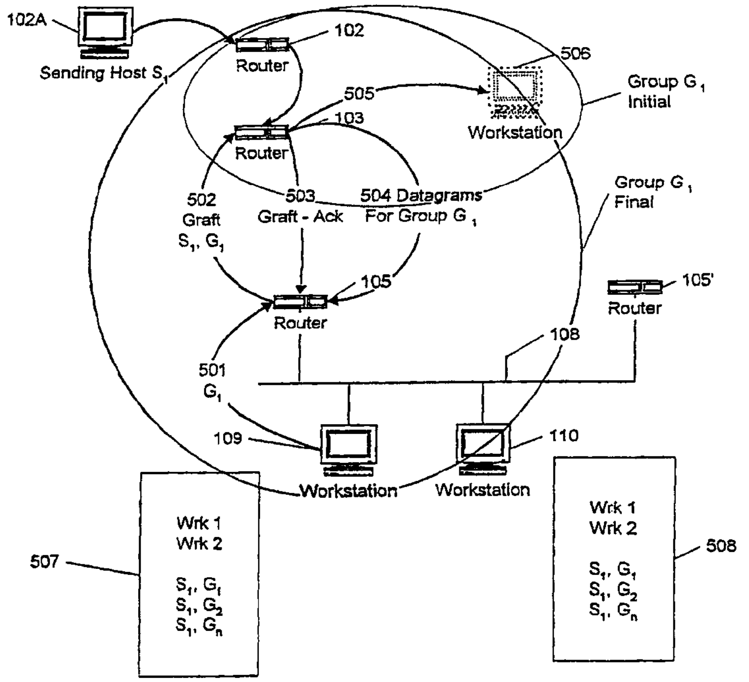
(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD, APPARATUS, AND MEDIUM FOR MINIMAL TIME MULTICAST GRAFT/JOIN RESTORATION

(57) Abstract

The present invention provides a method, apparatus, and medium for quickly re-establishing a lost multicast connection between an end user and a group in a multicast environment. The end user monitors the liveness of the received information as well as retains a list of the multicast communication channels required for re-establishment of a connection to a group. Through the use of an end user-originated Graft or Join based on the stored list of multicast communication channel identifiers (S, G), when the received information is no longer live, an end user may quickly rejoin a multicast group with minimal down time.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

Method, Apparatus, And Medium For Minimal Time Multicast Graft/Join Restoration

Background Of The Invention

1. Technical Field

The invention generally relates to IP multicast technology. More particularly, the invention relates to re-establishing a link with minimal delay, between a user's local area network and a multicast content channel.

2. Related Information

As the Internet becomes increasingly burdened with traffic, solutions for relieving at least some of the burden on the current Internet infrastructure are being sought. Current IP packet transmissions are sent point to point using point-to-point type protocols (PTP) as well known in the art. An example of point-to-point protocol is TCP/IP. Using PTP protocols to send data become increasingly inefficient in terms of bandwidth utilization as final destinations are located closer to each other. Packets traveling to similar destinations may traverse similar paths in a network, thus consuming extraneous bandwidth along the way. Also, the transmitting host faces a mounting burden as it attempts to service the numerous destinations on a timely basis. One solution is the use of IP multicast technology.

Also referred to as IP multicasting, this technology is a mechanism provided by a network that facilitates the transmission of a single packet to a plurality of destinations. When a network node (router or switch) handling the multicast packet in transit determines that the end destinations to which the packet is heading no longer uses a similar pathway, the network node duplicates and transmits the packet down as many pathways as needed to forward the packet to all predetermined locations specified by the multicast communication channel. The class of controlling protocols utilized by network nodes that build and maintain multicast communication channels (or routing trees) are referred to as IP multicast routing protocols. The class of controlling protocols utilized between destination machines and their first hop network nodes are called Group Management Protocols, such as IGMP. The IGMP specification (identified as Internet Group Management Protocol, Version 2, Network Working Group of the Internet Engineering Task Force, November 1997), as is known in the art, is hereby incorporated by reference.

Multiple IP multicast routing protocols exist, including, but not limited to, PIM dense mode, PIM sparse mode, CBT, and MOSPF. For example, PIM (Protocol Independent Multicast) dense mode is a multicast routing protocol that controls the maintenance of multicast communication channels utilized for transmissions to multicast groups which are densely distributed across a network. Applicable networks include intranets, extranets and the Internet, and other equivalent networks. PIM dense mode uses reverse path multicasting (RPM), also called reverse path forwarding (RPF), to establish and maintain routing trees. RPM is a technique in which a multicast packet (or datagram) is forwarded if the receiving interface on a network node is the one used to forward unicast datagrams to the source of the multicast datagram.

The first step in establishing a routing tree according to the PIM dense mode specification (Protocol Independent Multicast Version 2, Internet Engineering Task Force, April 2, 1997), hereinafter incorporated by reference, is to broadcast the multicast datagram to all PIM DM enabled routers in the network, such that a routing tree is formed that provides a multicast communication channel on which datagrams are carried. This broadcast process is repeatedly performed every three minutes. Figure 1 shows an example of setting up a routing tree in a PIM dense mode IP multicast system. Network 101 includes a plurality of network nodes 102-107. Also, Figure 1 includes LAN 108 as attached to network node 105 and workstations 109 and 110. As is well known in the art, the communication protocol between network node 105 and workstations 109 and 110 on LAN 108 may include the Internet Group Management Protocol, IGMP. In addition, other communication protocols are available. For example, there are several version of IGMP, mainly version numbers 1, 2 and 3, the last of which is a draft proposal in the research community. Other Management Protocols exist. For example, the Cisco Corporation has developed a protocol called CGMP (Cisco Group Management Protocol) which is similar to IGMP, described above. Here, IGMP is used to tell network node 105 (leaf node) that workstations 109 and 110 exist and continue to have interest in the particular multicast communication channel on which datagrams are forwarded to them. In particular, network node 102 receives a datagram from sending host 102A and transmits it to downstream network nodes 103, 104, and 107 as illustrated by transmission paths 111, 112, and 113. Having received the datagram, network nodes

103, 104 and 107 retransmit the datagram to other network nodes to which they are connected, except to the router from which the datagram was received. In this case, network node 103 transmits the datagram to network node 104 via path 114 and to network node 105 via path 115. Network node 104 retransmits the datagram to network node 103 via path 118, network node 105 via path 120, network node 106 via path 121, and network node 107 via path 119. Network node 107 retransmits the datagram to network node 104 via path 116 and network node 106 via path 122. Finally, depending on circumstances, for example, when the datagram was received by network node 106, it may retransmit the datagram to network node 107 via path 117.

The second step in establishing PIM DM routing trees includes pruning back all branches that lead to end stations that have not expressed interest in attaching to the multicast communication channel. Figure 2 shows an example of the pruning process in action in accordance with the PIM dense mode. In Figure 2, it is assumed that either of workstations 109 or 110 have expressed an interest in attaching to the multicast communication channel (that is currently in the broadcast phase, as shown in Figure 1). Router 105 sends prune message 203 to router 104, but does not send a prune message to router 103, because the RPF (the best reverse path to the source of the multicast data) interface on router 105 is the one on which router 103 is connected. Router 105 will not prune itself from the multicast channel in which it is interested. It is assumed, although not shown, that router 106 servers at least one end user workstation that has expressed interest in the multicast channel. As a result, router 106 does not prune its interface towards router 107, because the interface of router 106 knows that router 107 is router 106's RPF to the sending host 102A. Router 107, however, does send prune message 205 to router 104.

While Figure 2, as relating to PIM dense mode, shows a mechanism used to establish routing trees, other techniques to establish routing trees are known in the art as related to other IP multicasting protocols including protocols PIM sparse mode, CBT, and MOSPF, for example. In particular, PIM sparse mode is another applicable Multicast Routing Protocol (Protocol Independent Multicast-Sparse Mode: Protocol Specification, September 9, 1997), hereinafter incorporated by reference.

The resulting network as created by routing tables stored in the various network nodes is shown in Figure 3A. The resulting multicast channel is shown in Figure 3B. Datagrams, originating from sending host 102A, are transmitted from network node 102 to network nodes 103 and 107 via paths 301 and 302, respectively. Next, the datagrams are transmitted to network nodes 105 and 106, via paths 303 and 305, respectively. For a multicast routing communication channel to terminate at a network node, the network node preferably has at least one destination end user attached to it and that the destination end user has expressed an interest in attaching to the multicast communication channel (via IGMP or some similar Group Management Protocol). It is noted that, while multicast routing communication channels terminate at network nodes, multicast communication channels terminate at end user workstations. As shown in Figure 3A, there are no end users attached to network node 104. Accordingly, while it is possible to establish and maintain a communication path to network node 104 regarding a multicast channel, the lack of end users connected to network node 104 suggests that the connection for this channel should be pruned back. In an alternative embodiment of the invention, network node 104 may remain connected to receive a multicast channel so as to be available for quick connection to the multicast channel by another end user.

One feature of multicast systems includes the concept of groups. Groups are sets of destinations that participate in a multicast communication channel. The channel generally originates with a single content provider. However, multiple content providers may combine to provide content on the same multicast communication channel. Accordingly, to specify a content channel precisely, one needs to know the source of the information as well as the group identifier to which the channel is directed. Figure 4 shows network node 102 (of network 101, not shown for simplicity) receiving a multicast datagram from sending host 102A and sending the datagram via path 301 to network node 103. Network node 103 transmits the datagram to network nodes 403, 404, and 405. Originally, all the network nodes including 102, 103, 403, 404, 405, and 105 participated in group G_1, during the broadcast phase. After the pruning phase, only network nodes 102, 103, 403, 404, and 405, remain participating in G_1 (105 having pruned back the channel).

Next, if workstation 109 (or 110) which is connected to network node 105, part of network 101, desires to join group G_1 409, it sends a request via IGMP to network node 105 specifying that it wishes to join group G_1. Network node 105 is considered to be a leaf node, in that it is directly connected via a LAN to end users that have expressed interest in the particular multicast channel, and there are no further routers downstream of router 105 on the LAN. Network node 105 interprets the request from workstation 109 as a command to attempt to join group G_1. Network node 105 sends a Graft message with the payload G_1, S_1 to network node 103 via path 401. In response, network node 103 replies with a Graft acknowledge (Graft-Ack) signal to network node 105 via path 402. When network node 103 receives the next datagram destined for group G_1 409, it transmits the datagram to network node 105 as network node 105 is now part of group G_1 409.

It is an assumption in the above that the state $G_{1,}$ S_1 is retained indefinitely, in network node 105 from the previous broadcast phase. However, it is possible that the state may not exist in network node 105, in which case the network node 105 cannot send a Graft towards network node 103 because network node 105 does not know from which network node it can obtain the specific multicast channel.

Further, in the above-described system, as well as in other IP multicast systems, if a leaf network node fails, the system may have to wait between 0 and 3 minutes (the PIM dense mode broadcast phase cycle time) or for an average of 1.5 minutes until the multicast system enters the broadcast phase such that state is re-established in network nodes and the channel can re-establish itself. A difficulty here, especially in environments with a need for minimal disconnect time, is that the delay associated with re-establishing a connection to network node 105, if it becomes separated from group G_1, may take an average of 1.5 minutes. The cause of the separation may be due to a variety of reasons including transmission line failure, local router failure, and related problems. This reconnect delay is the average time in which

PIM dense mode enters the re-broadcasting phase as shown in Figure 1. In systems that require faster reconnect times, for example, in financial arenas where the 1.5 minute average delay freezes a trader's information stream so the trader cannot act as desired, the delay associated with waiting until the overall multicast system re-establishes itself is unacceptable. Further, merely decreasing the multicast system's re-broadcast interval substantially increases the amount of non-usable broadcast information sent to all end network nodes, regardless of their interest in the particular communication channel. This approach can be very costly in terms of inefficient use of network bandwidth. The result of this delay may be one reason that systems requiring a fast re-establishment time (high availability requirements) have not readily embraced multicast technology as a solution for efficient information transfer.

In networks with only one leaf router serving a LAN, if that leaf router fails, connectivity to the group is lost. For example, if network node 105 becomes disabled, LAN 108 is prevented from connecting to group G_1. One possible solution is to utilize a more redundant system, in which there are multiple leaf network nodes. For example, one could add another network node 105' to the LAN 108 to the network, by connecting to network node 103 through path 306 as shown in Figure 3A. In the above-described environment, one of these network nodes 105 and 105' (for example, 105') will have been pruned out of the multicast channel during the pruning phase of the PIM DM protocol, resulting in the multicast channel as shown in Figure 3B.

The dotted line to network node 105' in Figure 3A indicates that while 105 and 105' are considered separate network nodes, as part of the PIM DM protocol, one of them (i.e., 105') has pruned itself from the multicast channel 101, leaving the remaining network node (105) as the "forwarder" as represented in Figure 3B. The process to elect which network node is to be the forwarder to LAN 108 is called Assertion. Routers 105 and 105' execute the Assert process, as specified in the PIM dense mode specification, and only one of the network nodes becomes the forwarder, and the other network node(s) on the LAN that lose the assertion process prune themselves back, e.g. 105'. Accordingly, when network node 105' has pruned itself from the multicast channel, it needs to wait for the re-broadcast phase (see Figure 1 implemented every three minutes) before network node 105' can begin receiving new datagrams again (see Figure 3A). When this occurs, the assertion process is repeated and the network nodes that lose the assertion process prune themselves back.

The existence of two client side network nodes (leaf nodes in this case) 105 and 105' is to make LAN 108 more robust in being able to handle router or switch failures (network node). However, at most only one of the two network nodes 105 and 105' is active in receiving datagrams from network node 103 and forwarding them onto the LAN 108, at any given time, because the other node (e.g., 105') was pruned out of the multicast channel during the pruning phase. Note that during the assertion process, both network nodes 105 and 105' can be acting as forwarder to LAN 108, until one of the network nodes wins the election process and becomes the forwarder for the LAN 108. If network node 105 is the forwarded for the LAN 108, and in the event that network node 105 becomes disabled, workstations 109 and/or 110 on LAN 108 will have to wait until the PIM DM re-broadcast phase initiates, so as to forward packets to network node 105' (an average of 1.5 minutes). When 105' begins to forward packets onto LAN 108, network node 105' will elect itself as the forwarder (since no other network node has challenged it and the assert election process is not triggered in this case) and being to forward packets to onto the LAN 108. While waiting for reconnection to the multicast channel, the hosts on LAN 108 may experience large gaps in messages, and general data loss. Again, the maximum waiting interval is, as above, 3 minutes, with an average waiting interval of 1.5 minutes. In applications that require high availability, with or without high throughput characteristics, the average waiting interval of 1.5 minutes is unacceptable as volumes of data may back up, resulting in excessive data loss, memory usage and delay in attempting to re-establish the data to the hosts on LAN 108. Furthermore, users of real-time applications that require up-to-the-second information cannot tolerate even modest periods of disconnection from the service, e.g. instrument traders using a real-time financial application.

Summary Of The Invention

The present invention overcomes the above-described problems by providing a fast re-establishment of a lost connection to a multicast group. To quickly reconnect an end user (or an end user's local area network, or LAN), the end user's system stores relevant connection information so that, when needed, it directs a network node to re-establish its connection to one or more multicast groups.

Each LAN end user on network 108 may elect a director for the LAN 108 that is responsible for coordinating the recovery of the one or more multicast channels to the LAN. Alternatively, a director may not be elected and each end user may act independently and redundantly perform the operations that the director would handle. Advantages of including a director include limiting the task of initiating re-establishment of the multicast channel to one designated entity, rather than several end user workstations. Advantages of not including a director include redundant processing so that if any portion of the LAN 108 fails, each workstation (109, 110) may be able to re-establish its connection with network 101. To this end, when referencing the director, it will be understood that the director is intended to refer to any workstation performing the functions of the director, including but not limited to any end user's workstation redundantly performing the director's functions where LAN 108 has no director and each end user's workstation acts independently.

The director stores the group information as received from various sources. For example, the director may store this information in a source, group pair (also referred to as an S, G pair). The director also monitors sender "liveness" regarding a monitored group. Liveness, for purposes herein, refers to the monitoring of information received from a source. By monitoring liveness information, the receiver (director or workstation) knows when to expect information from the source. That is, when the source is idle (has no data to send), liveness (heartbeat) messages are sent at predefined intervals or at decaying intervals to inform all end user systems that the channel is still alive. Various sender driven liveness messages are specified in

Holbrook et al.'s "Log-Based Receiver" (H. Holbrook, S. Signhal, D. Cheriton: Log-Based Receiver-Reliable Multicast for Distributed Interactive Simulation. Computer Communication Review, Vol. 25, No. 4, Proceedings of the ACM SIGCOMM'95, August 1995) and are incorporated herein by reference.

Once the information has not been received as expected, the director determines that the connection to the source has been lost. Next, the director retrieves the previously stored information (the S, G pair), forms a Graft message, and transmits a Graft message to the all routers address (224.0.0.13) or to a specific leaf router address, if the target leaf router address is well known (configured or learned) by the director. The all routers address is specified in the PIM DM specification. Note that the Graft message could be sent using any defined or reserved address that delivers the message to the LAN network nodes that can process the Graft and re-establish the multicast channel to the LAN. The network node or network nodes receiving the Graft message acts accordingly to re-establish the connection to the group by immediately forwarding the Graft upstream on the Reverse Path Forwarding (RPF) interface, as specified in the PIM DM specification. In this scenario, the director is acting like a network node by forwarding a Graft to its leaf router(s). The receiving network node does not distinguish between the director and another node, in this case, and processes the Graft as if it were received from a network node, as specified in the PIM dense mode specification. The result is that the multicast channel is more rapidly re-established the LAN. Thus, by using the system as embodied by the invention, a director may quickly re-establish a connection to a group with minimal delay.

As purposes herein, the Internet is used as an example of a network of computers which benefits from IP multicast technology. However, it will be understood that the invention as described herein may be readily applied to internets, extranets, WANs and other networks which may benefit from multicasting. Further, while the invention is described in connection to PIM dense mode, it is readily apparent that the advantages described herein are applicable to other IP multicasting protocols including PIM sparse mode, CBT, and MOSPF, and other equivalent IP multicasting protocols. Accordingly, the reference to PIM dense mode is made by way of example only. For example, in PIM sparse mode, the director would send a Join message rather than a Graft, when the liveness information has not been received as expected. While PIM sparse mode is not a broadcast and prune protocol, as is PIM dense mode, the rapid recovery mechanisms described herein are applicable in achieving rapid multicast channel recovery.

Brief Description Of The Drawings

In the following text and drawings, wherein similar reference numerals denote similar elements throughout the several views thereof, the present invention is explained with reference to illustrative embodiments.

Figure 1 is network diagram showing a broadcast phase of a conventional multicast system.

Figure 2 is network diagram showing a pruning phase of a conventional multicast system.

Figures 3A and 3B are network diagrams showing a remaining network after a broadcast and pruning phases of a conventional multicast system.

Figure 4 is a network diagram showing a Graft in a conventional multicast system.

Figure 5 is a network diagram of a Graft originating with an end user according to embodiments of the present invention.

Figure 6 is a packet format diagram of a multicast system as used in conjunction with embodiments of the present invention.

Figure 7 is a network diagram of the present invention including an embodiment of an end user's system according to embodiments of the present invention.

Figure 8 is a network diagram of the present invention showing Graft messages as contemplated by the present invention.

Detailed Description Of The Preferred Embodiments

The present invention relates to quickly re-establishing a connection to a multicast group. For purposes herein, a multicast group should be understood to also relate to a multicast communication channel.

Figure 5 shows an array of workstations 109 and 110 on network 108 (for example, a LAN) connected via network node 105 to network 101 (not shown for simplicity). While only two workstations are shown, it is understood that many workstations may be connected to network 108. One of the workstations, 109 or 110 may be elected as director for a particular S,G pair. As the director officiates in numerous occasions, the director may be also referred to as a host of the network 108. In this example, workstation 109 is the elected director. In case director workstation 109 fails, workstation 110 is the designated assistant director. The assistant director stores information similar to workstation 109 so as to replace workstation 109 in the event workstation 109 fails.

To this end, when referencing the director, it will be understood that the director is intended to refer to any workstation performing the functions of the director, including but not limited to any end user's workstation redundantly performing the director's functions where LAN 108 has no director and each end user's workstation acts independently.

Workstation 109 stores information 507 in internal memory. While not shown for simplicity, workstations 109 and 110 may contain various forms of internal memory including RAM, ROM, replaceable storage devices (for example, diskettes, hard drives, and CD-ROMs), and equivalents thereof. Information 507 contains a list of which workstations are the director and assistant director (wrk1 and wrk2 respectively) and a list of S, G pairs that list the source (for example, S_1) of various groups (for example, G_1, G_2, through G_n). Workstation 110 may store a similar set of information 508.

The director derives sender liveness information from information received in connection with the received liveness datagrams, the latter (liveness or heartbeat mechanisms) being discussed in the Holbrook et al. "Log-Based Receiver". From this sender liveness information, the director determines when to expect datagrams from a source S regarding group G. Once the datagrams or liveness messages are not received as expected, the director transmits a Graft or Join message 501 in PIM DM or PIM SM, respectively, with the appropriate payload (for example, S_n, G_1) to the all routers address (as found in the PIM DM specification). Alternatively, the director transmits the Graft message 501 to a specific leaf router, or any other all local routers address. If the director did not receive the liveness messages as expected, because the forwarder for the multicast channel, e.g. router 105, failed, then the other router, e.g. 105' (or routers) will receive the Graft message and forward it to network node 103 (or on their respective RPF interface), which will cause the quick re-establishment of the multicast channel to LAN 108. Assistant director workstation 110, in addition to being able to perform tasks similar to that of director workstation 109, monitors director workstation 109 in its performance of its director duties. So, if and when director workstation 109 fails, workstation 110 may act as a backup. As noted above, if the PIM sparse mode multicast routing protocol was used instead of PIM dense mode, then the director would send a Join message instead of a Graft. For purposes of simplicity, a workstation and equivalent systems may be referred to as information handlers.

The director generates the S,G pair(s) by monitoring received datagrams for the respective multicast channel(s). The header information in a received IP Multicast packet contains the sender's identity (here the identify of the source end user's workstation, and the group address G, carried in the IP destination field of the IP header. The workstation stores the sender's information in conjunction with the group over which the datagram was received. For example, in a UDP header, the sender transmits the following: a source address, a destination address, the source port, and the destination port. In a multicast environment, the source address is the IP address of the host sending the information out to receivers. For purposes herein, this source address is the S of the S, G pair. The destination address is the IP address of the multicast communication channel referred to as a group. This group is the group G of the S, G pair. Also, end user workstations can further refine that information that is passed to the end user application for a given G, by filtering those datagrams that are received on destination port numbers of which are interest to the receiving application(s). Workstations 109 and 110 would derive the S, G pair from datagrams received from an ongoing multicast channel via network node 105.

When workstation 109 desires to connect to a multicast channel G_1 (including routers 102 and 103 and workstations 506), it requests connection to this channel G_1 via IGMP as described above. At this point, network node 105 determines which sources (S_1, S_2, etc.) correspond to the requested channel. Next, network node 105 forwards all channels to LAN 108 which correspond to the specified G. So, if S_1,G_1 and S_2,G_1 multicast channel exists in the network, then network node 105 forwards both S_1,G_1 and S_2,G_1 to workstation 109. The application layer in workstation 109 is then responsible for sorting out which, if any, of S_1,G_1 and S_2,G_1 is the actual channel desired. It is possible that the application may desire to obtain all of the S, G pairs.

After workstations 109 and 110, join group G_1, group G_1 is understood to include routers 102, 103, and 105 and workstations 506, 109, and 110.

Figure 6 shows an example of PIM packet formats for control messages as sent from workstation 109 and 110. Bits 0-3 relate to the PIM version, bits 4-7 identify the type of control message, bits 8-15 are reserved, and bits 16-31 are checksum. The following list defines PIM message types:

0 = Hello

1 = Register

2= Register - Stop

3 = Join/Prune

4 = Bootstrap

5 = Assert

6 = Graft

7 = Graft - Ack

8 = Candidate-RP-Advertisement

Here, the Graft control message as part of the present invention is initially formatted in workstation 109 or workstation 110 depending on which workstation(s) is attempting to re-establishing the link to group G. The other control messages can be found in the PIM dense mode or PIM spare mode specification, referenced here in, etc.

Figure 7 shows an embodiment where more than one router is connected to LAN 108. In Figure 7, two routers 701 and 702 connect LAN 108 to router 105. The interaction between routers (network nodes) 701 and 702 and router (network node) 105 and workstations 109 and 110 is as described above. In some applications, workstation 109 may be the only workstation. In this case, LAN 108 may be supporting a single workstation 109 in connection to network 101. To this end, the term LAN may refer to a device which allows workstation to connect to both network nodes 701 and 702 simultaneously.

Another embodiment of the system described herein is shown with respect to Figure 9. Figure 9 shows network node 105, network nodes 701 and 702, and LAN 108 as described with respect to Figure 7. Also included in the arrangement are network nodes 703 and 704 also connected to network node 105 and LAN 705. Next, both LANs 108 and 705 are connected to workstation 109. In this embodiment, both LANs 108 and 705 support a single (or multiple) workstations. By using this arrangement, workstation 109 (and other similarly connected workstations) is provided with multiple paths to connect to network node 105. An example of when the arrangement Figure 9 may be used is in applications where redundant paths to the network 101 is required, so as to achieve higher availability of services provided by sources such as 706.

Figure 8 shows a network level diagram of network 801 as including a variety of sources S_n and a variety of end user workstations 810-812. Source S_1 transmits group G_1 datagrams to receivers S_4 and S_5 (S_4 and S_5 are receivers with respect to group G_1, but are senders with respect to the groups they source) on multicast channel S_1 G_1. Notably, the combination of receivers S_4 and S_5 comprise group G_1.

Also, source S_2 sends content on group G_1 (on multicast channel $S_2 G_1$) as shown by the combined use of G_1 from sources S_1 and S_2. Alternatively, the content of G_1 may originate with S_1 and use S_2 as a backup for content. Source S_1 also provides information to receiver S_3. Having received datagrams for group G_2, receiver S_3 reformulates and retransmits the datagrams to receivers of network 809. Network 809 may be a LAN or individual groups of receivers so long as they maybe collectively referred to as group G_5. Similarly, receiver S_4 reformulates and retransmits received datagrams to network 808 as group G_3. Finally, receiver S_5 reformulates and retransmits received datagrams to network 807 as group G_4.

When receiver 810 (for example, a workstation with the functionality of workstation 109 of Figure 5) from network 809 wishes to re-establish a connection to group G_5, receiver 810 transmits a Graft message as described above to its leaf router containing the S_3,G_5 pair information. The network node (not shown) in between S_3 and LAN 809 receives G (here, G_5) and sends receiver 810 the multicast channel corresponding to S_3,G_5. In one embodiment, the Graft sent by the director specifically only re-establishes the specific S,G pair and not all *,G pairs, as is done with simply a IGMP join message from the host. Alternatively, all *,G pairs are re-established then unwanted ones dropped.

Similar operations from workstations 811 and 812, when they generate to their respective leaf routers Graft messages specifying the S_4, G_3 pair and the S_5, G_4 pair, respectively.

It should be noted that the relationships as described between S and G for Figure 8 are relative to the datagrams originating with sources S_1 and S_2. When a network receives information from alternative sources, for example, network 808 receiving group G_n datagrams (not shown for simplicity) from source S_3, the Graft S, G pair information from receiver 811 takes the form of (S_3, G_n).

It is readily understood that the method described above may be implemented through the use of a computer-readable medium. For example, one may store computer-implemented steps in a medium which is then read by any of workstations 109 and 110. Further, the process steps may be transferred into an internal storage of workstations 109 and 110 (or into alternative network nodes, for instance) through a different input into LAN 108 or network 101.

Using the above-described system, delay times in re-establishing a connection to a group may be minimized through sensing a fault in received (or not received) information, determining how to quickly reconnect, and reconnecting to a group through the use of information stored in a monitoring station.

It will be apparent to those skilled in the art that application of the present invention need not be utilized in conjunction with the Internet. It is also envisioned that the techniques of the present invention apply to any network including intranets and extranets.

Further, the present invention may also be implemented in a peer-to-peer computing environment or in a multi-user host system having a mainframe or a minicomputer. Thus, the computer network in which the invention is implemented should be broadly construed to include any multicast computer network from which a client can retrieve a channel in a multicast environment.

In the foregoing specification, the present invention has been described with reference to specific exemplary embodiments thereof. In particular, reference has been made to the PIM dense mode specification. Although the invention has been described in terms of a preferred embodiment, those skilled in the art will recognize that various modifications, embodiments or variations of the invention can be practiced within the spirit and scope of the invention as set forth in the appended claims. Some variations include the use of PIM sparse mode, CBT, and MOSPF IP multicast protocols and other multicast protocols. All are considered within the sphere, spirit, and scope of the invention. The specification and drawings are, therefore, to be regarded in an illustrated rather than restrictive sense. Accordingly, it is not intended that the invention be limited except as may be necessary in view of the appended claims.

CLAIMS:

I claim:

1. A system for re-establishing a connection to a multicast channel comprising:

a first network node receiving multicast data from a network;

a second network node connected to said network;

an information handler connected to said first network node and said second network node, said information handler including

a memory for storing reconnect information, and

one or more monitoring device(s) that monitors the reception of said data for a disconnect of said network node from said multicast channel and alerts said information handler upon occurrence or detection of said disconnect,

wherein, upon reception of said alert, said information handler retrieves said reconnect information and transmits said information to at least said second network node so as to reconnect said information handler to said multicast channel.

2. The system according to claim 1, wherein said information handler transmits said information to a reserved address specifying all routers.

3. The system according to claim 1, wherein the transmission of said information is in the form of a well formed packet.

4. The system according to claim 3, wherein the well formed packet is in the form of a PIM dense mode Graft.

5. The system according to claim 3, wherein the well formed packet is in the form of a PIM spare mode Join.

6. The system according to claim 1, wherein said reconnect information includes the source of said data and the group to which said data belong.

7. The system according to claim 1, wherein said information handler is a workstation.

8. The system according to claim 1, wherein said network routing protocol is based on the PIM dense mode specification.

9. The system according to claim 1, wherein said network routing protocol is based on the PIM sparse mode specification.

10. The system according to claim 1, wherein said first network node and said second network node are the same node.

11. A method for re-establishing a connection to a multicast channel comprising the steps of:

receiving in a first network node multicast data from a network;

storing in a memory in an information handler, which is connected to said first network node, reconnect information related to establishing a reconnection to said multicast channel;

monitoring the reception of said data for a disconnect of said network node from said multicast channel;

alerting said information handler upon occurrence or detection of said disconnect;

retrieving said reconnect information by said information handler upon reception of said alert; and,

transmitting said information to at least one of said first network node and a second network node to reconnect at least one of said first network node and said second network node to said multicast channel.

12. The method according to claim 11, wherein said transmitting step transmits said reconnect information to a reserved address specifying all routers.

13. The method according to claim 11, wherein said transmitting step transmits said reconnect information in the form of a well formed packet.

14. The method according to claim 13, wherein the well formed packet is in the form of a PIM dense mode Graft.

15. The system according to claim 13, wherein the well formed packet is in the form of a PIM spare mode Join.

16. The method according to claim 11, wherein said reconnect information includes the source of said data and the group to which said data belong.

17. The method according to claim 11, wherein said information handler is a workstation.

18. The method according to claim 11, wherein said network routing protocol is based on the PIM dense mode specification.

19. The method according to claim 11, wherein said network routing protocol is based on the PIM spare mode specification.

20. A computer-readable medium having computer-executable instructions for performing the steps comprising:

receiving at a first network node multicast data related to a multicast channel from a network;

storing reconnect information in a memory in an information handler, which is connected to said first network node;

monitoring the reception of said data for a disconnect of said first network node from said multicast channel;

alerting said information handler upon occurrence or detection of said disconnect;

retrieving said reconnect information by said information handler upon reception of said alert ; and,

transmitting said reconnect information to at least one of said first network node and a second network node to reconnect at least one of said first network node and said second network node to said multicast channel.

21. The computer-readable medium according to claim 20, wherein said transmitting step transmits said reconnect information to a reserved address specifying all routers.

22. The computer-readable medium according to claim 20, wherein said transmitting step transmits said reconnect information in the form of a well formed packet.

23. The computer-readable medium according to claim 22, wherein the well formed packet is in the form of a PIM dense mode Graft.

24. The computer-readable medium according to claim 22, wherein the well formed packet is in the form of a PIM sparse mode Join.

25. The computer-readable medium according to claim 20, wherein said reconnect information includes the source of said data and the group to which said data belong.

26. The computer-readable medium according to claim 20, wherein said information handler is a network node.

27. The computer-readable medium according to claim 20, wherein said information handler is a workstation.

28. The computer-readable medium according to claim 20, wherein said network is based on the PIM dense mode specification.

29. The computer-readable medium according to claim 20, wherein said network is based on the PIM sparse mode specification.

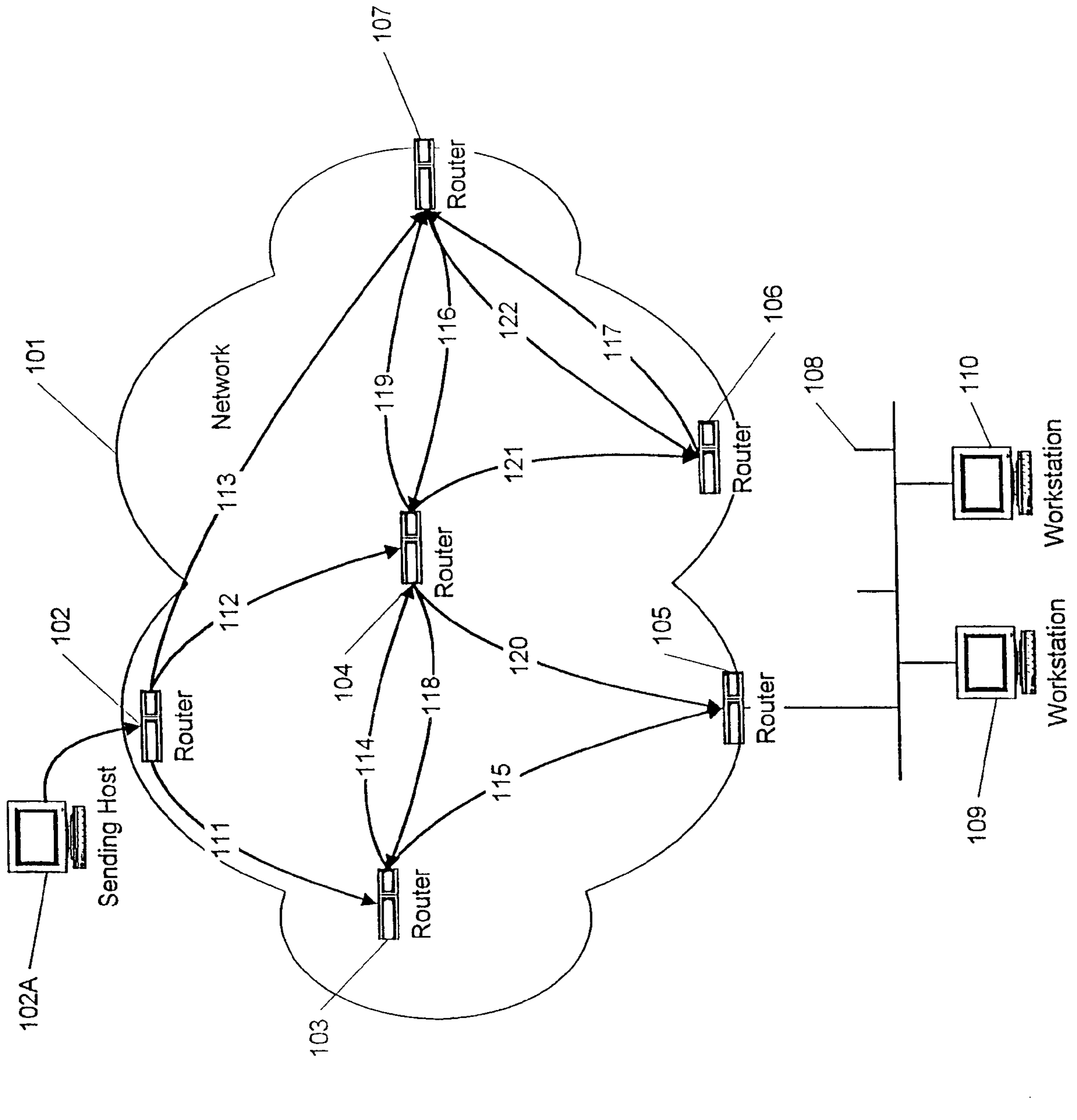

107
Router
101
Network
106
Router
116
119
122
117
113
121
Router
108
110
Workstation
102
112
104
Router
105
120
118
114
Router
Workstation
Sending Host
Router
111
115
109
102A
Router
103

FIG. 1

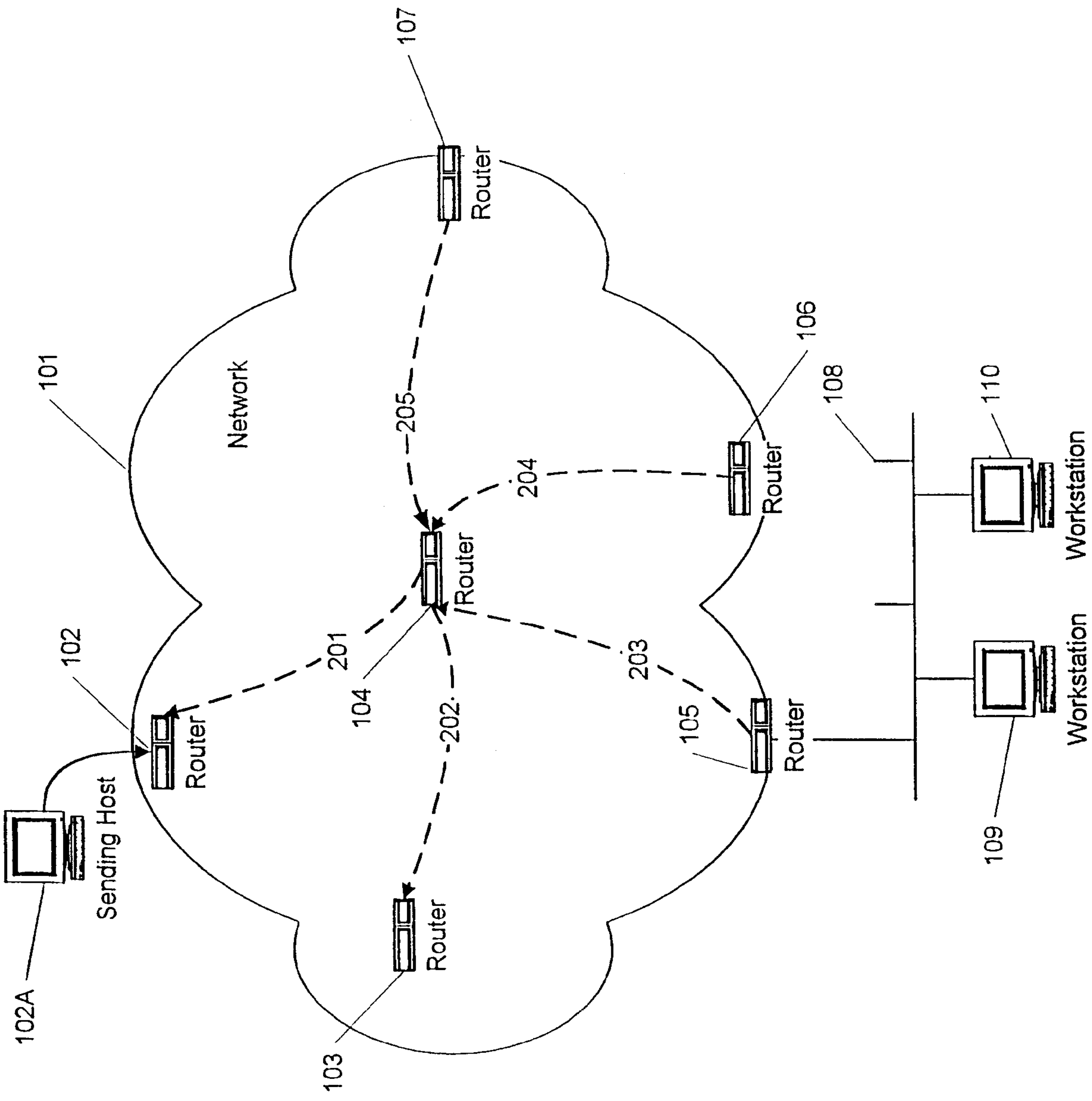

107
Router
101
Network
205
106
Router
204
108
110
Workstation
Router
104
201
102
Router
Sending Host
102A
202
103
Router
203
105
Router
Workstation
109

FIG. 2

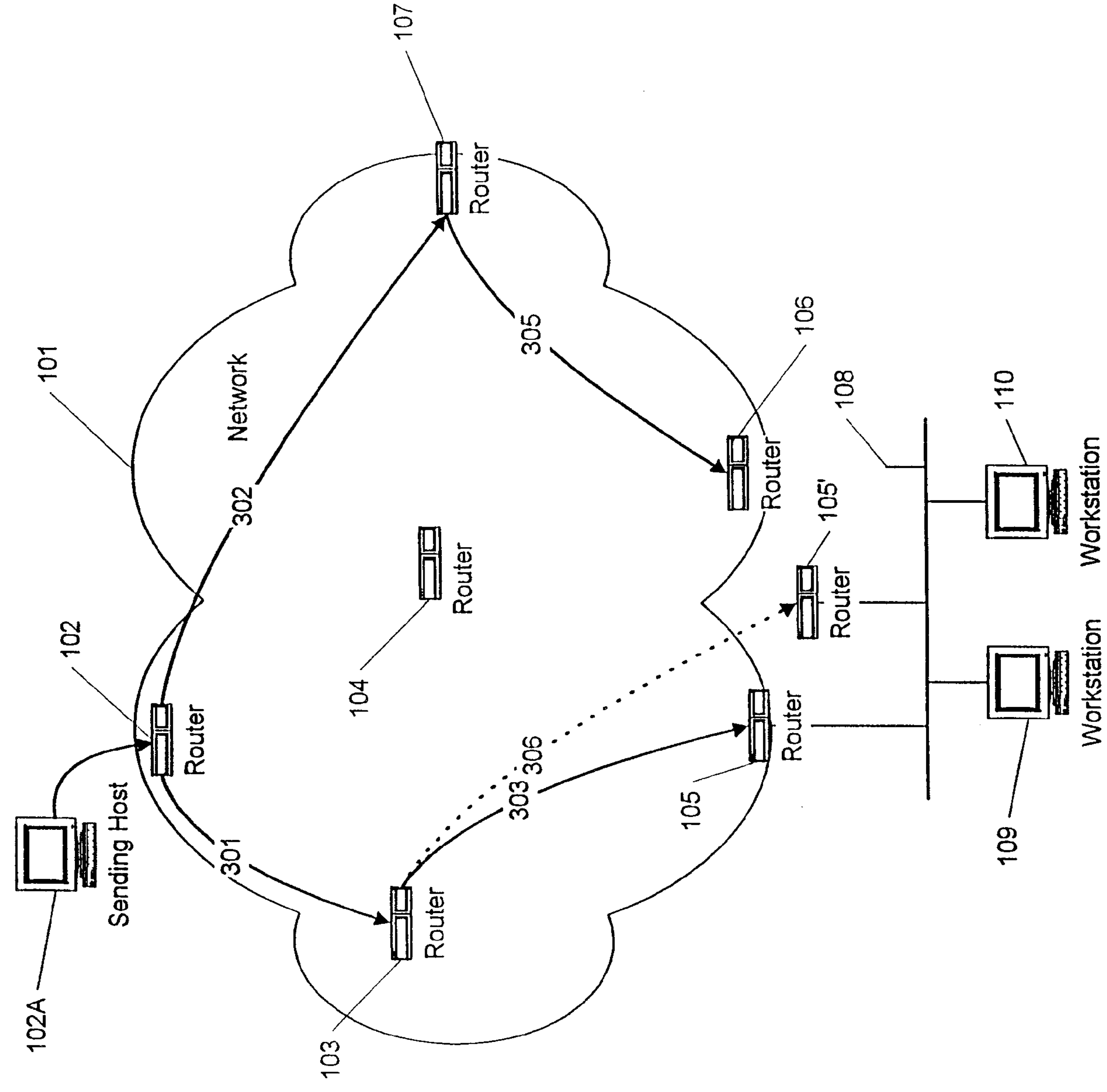

101
Network
102A
Sending Host
102
Router
301
302
103
Router
104
Router
303
306
105
Router
105'
Router
107
Router
305
106
Router
108
109
Workstation
110
Workstation

FIG. 3A

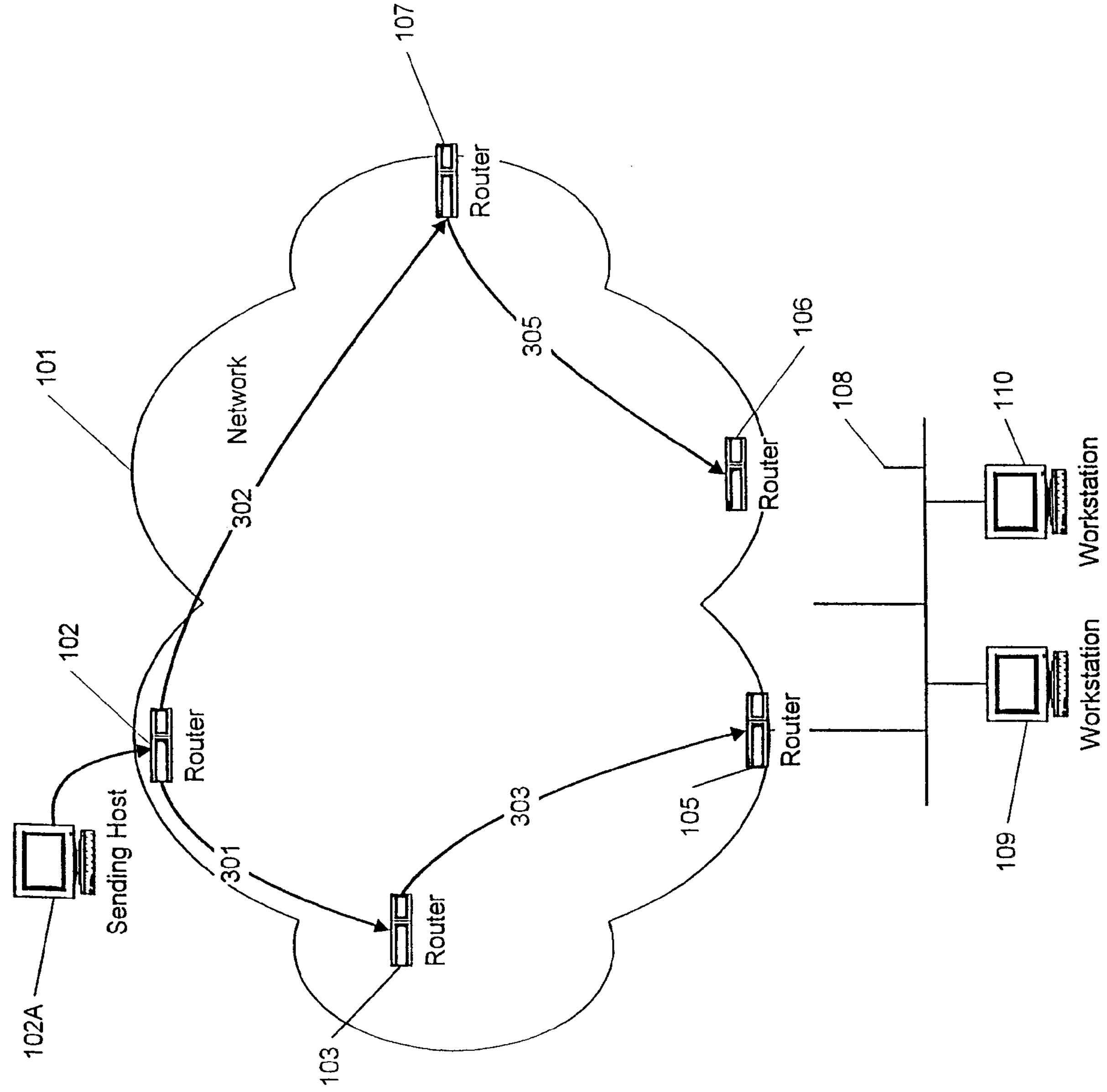
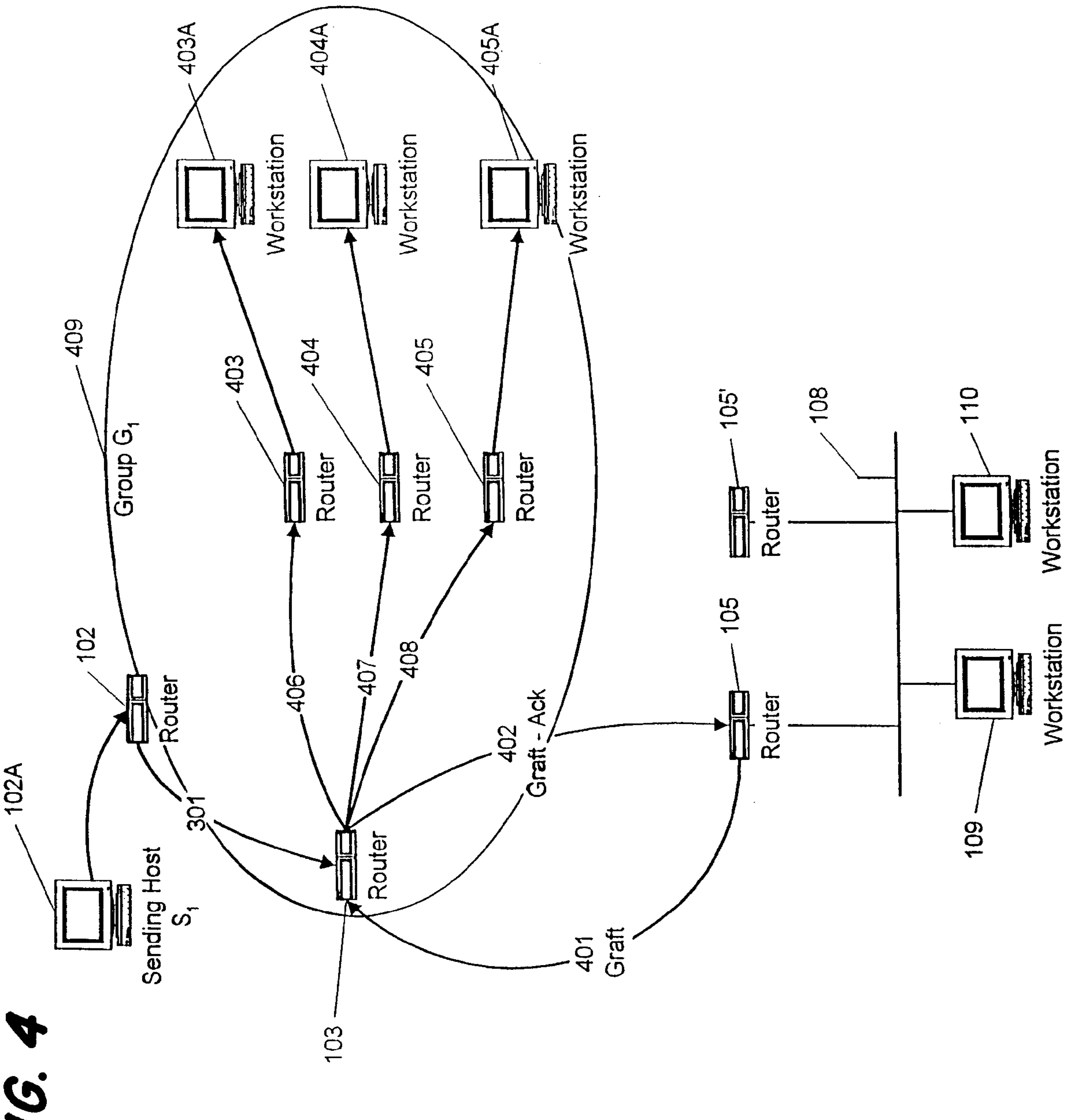

101
Network
102
Router
102A
Sending Host
103
Router
105
Router
106
Router
107
Router
108
109
Workstation
110
Workstation
301
302
303
305

FIG. 3B

FIG. 4


Sending Host S$_1$
102A
Router
102
301
Router
103
Group G$_1$
409
406
407
408
Router
403
Router
404
Router
405
Workstation
403A
Workstation
404A
Workstation
405A
402
Graft - Ack
401
Graft
Router
105
Router
105'
108
Workstation
109
Workstation
110

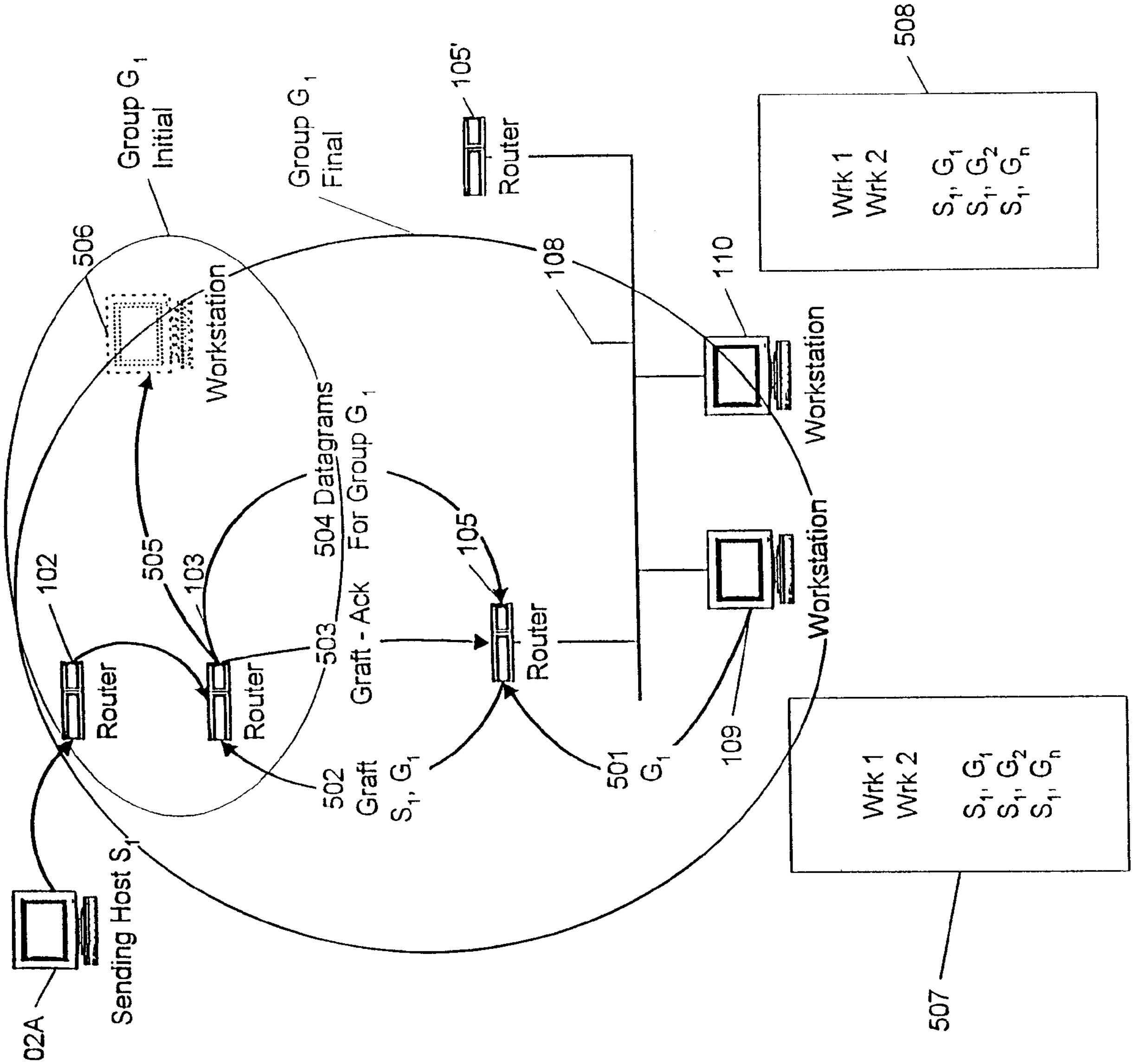

Group G_1 Initial
Group G_1 Final
105'
Router
508
Wrk 1
Wrk 2
S_1, G_1
S_1, G_2
S_1, G_n
506
Workstation
108
110
Workstation
504 Datagrams For Group G_1
102
505
103
105
Router
Router
Router
503
Graft - Ack
Workstation
502
Graft
S_1, G_1
501
G_1
109
Wrk 1
Wrk 2
S_1, G_1
S_1, G_2
S_1, G_n
Sending Host S_1
102A
507

FIG. 5

FIG. 6

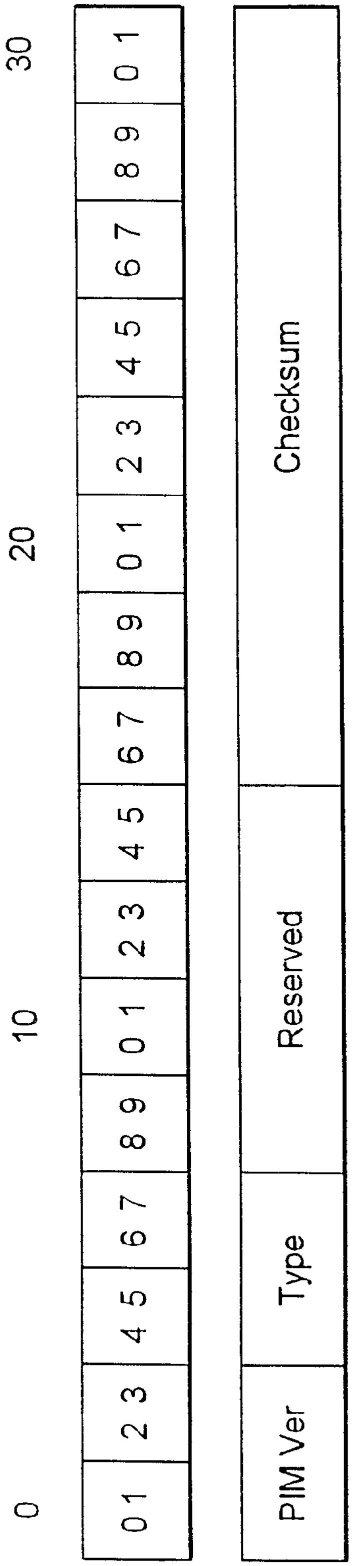

0
10
20
30
0 1
2 3
4 5
6 7
8 9
0 1
2 3
4 5
6 7
8 9
0 1
2 3
4 5
6 7
8 9
0 1
PIM Ver
Type
Reserved
Checksum

FIG. 7

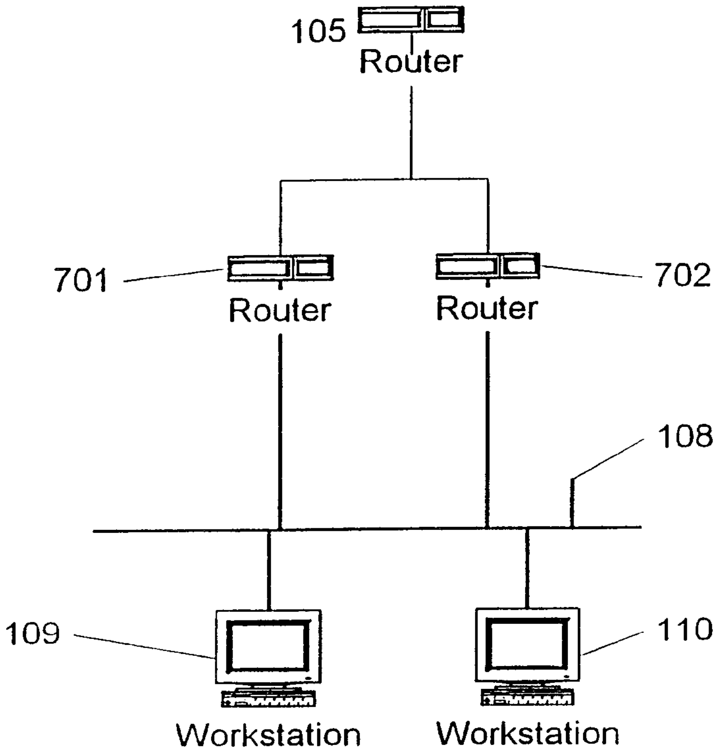

105
Router
701
Router
Router
702
108
109
Workstation
Workstation
110

FIG. 8

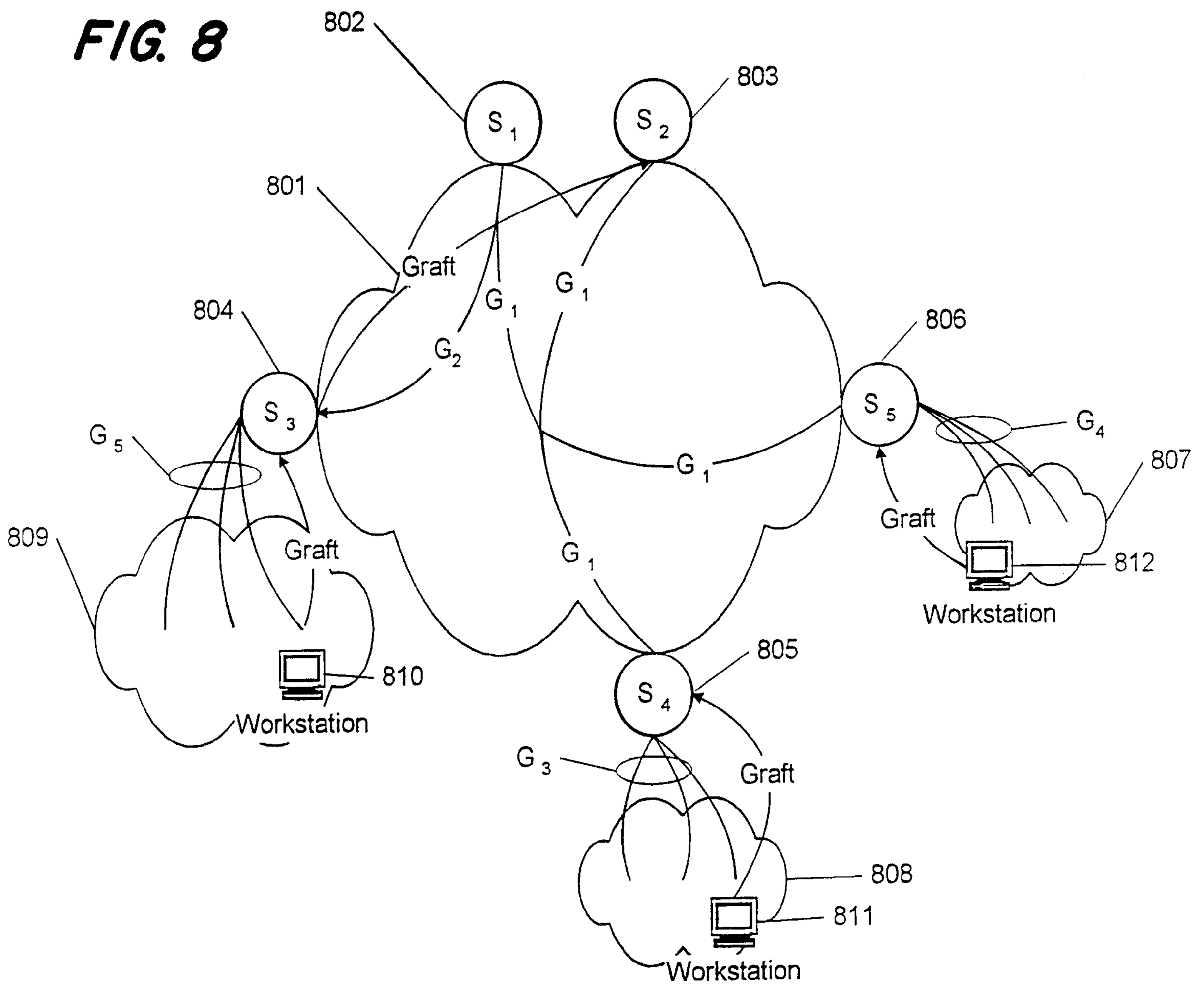
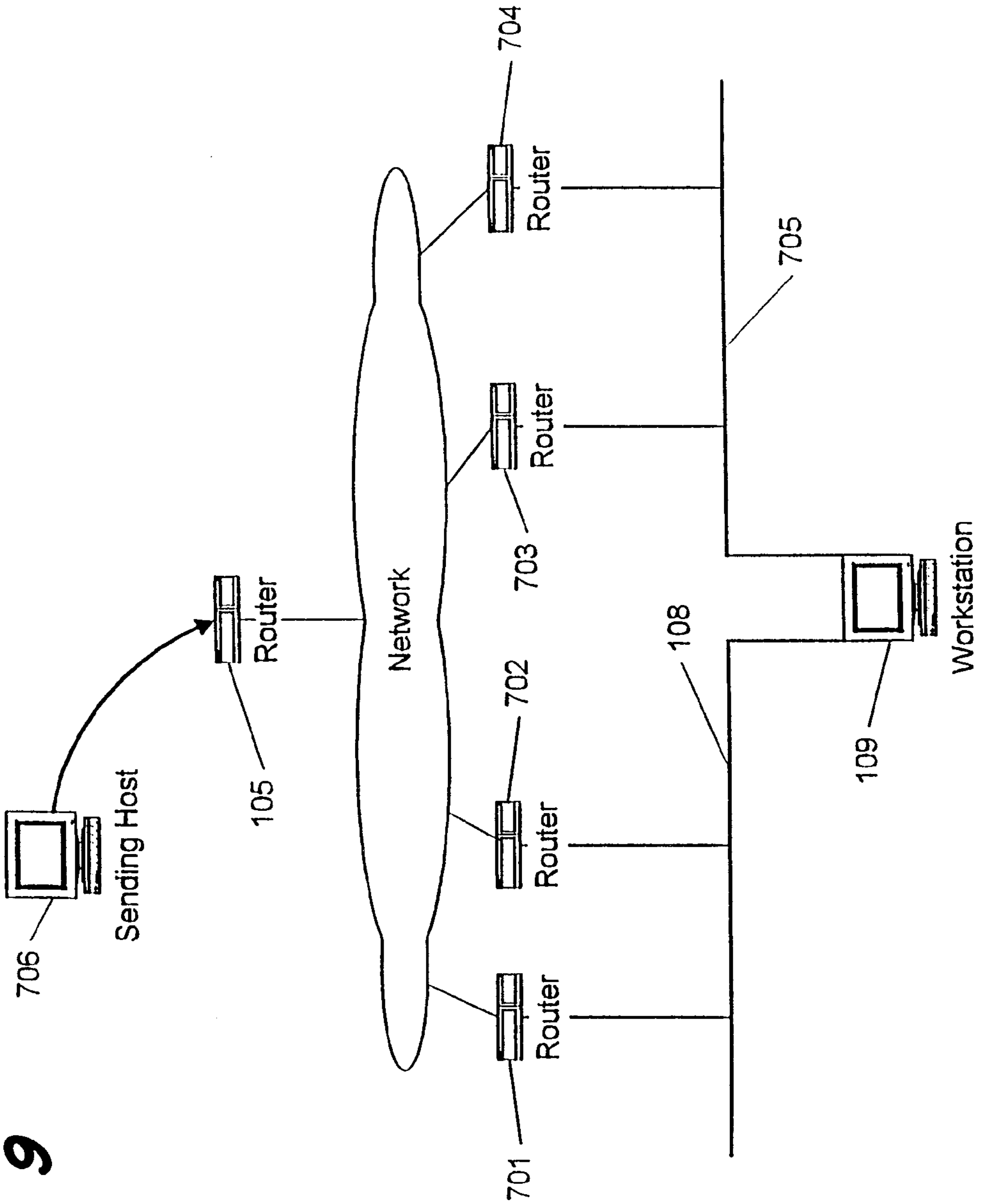

802
S 1
803
S 2
801
Graft
G 1
G 1
804
G2
S 3
806
S 5
G4
807
G 5
G 1
809
Graft
Graft
812
G 1
Workstation
805
810
S 4
Workstation
G 3
Graft
808
811
Workstation

FIG. 9


706
Sending Host
105
Router
Network
701
Router
702
Router
703
Router
704
Router
705
108
109
Workstation

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/US 99/05731

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 H04L12/18 H04L29/14

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	DEERING S ET AL: "AN ARCHITECTURE FOR WIDE-AREA MULTICAST ROUTING" COMPUTER COMMUNICATIONS REVIEW, vol. 24, no. 4, 1 October 1994 (1994-10-01), pages 126-135, XP000477046 ISSN: 0146-4833 page 129, right-hand column, line 14 - page 131, right-hand column, line 17 --- -/--	1-29

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search 19 August 1999	Date of mailing of the international search report 30/08/1999
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer RAMIREZ DE AREL.., F

2

INTERNATIONAL SEARCH REPORT

Inter onal Application No
PCT/US 99/05731

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	BILLHARTZ T ET AL: "PERFORMANCE AND RESOURCE COST COMPARISONS FOR THE CBT AND PIM MULTICAST ROUTING PROTOCOLS" IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, vol. 15, no. 3, 1 April 1997 (1997-04-01), pages 304-314, XP000683934 ISSN: 0733-8716 page 304, right-hand column, line 27 - page 306, right-hand column, line 4 page 312, right-hand column, line 38 - page 313, left-hand column, line 24 ---	1-29
A	EP 0 688 121 A (TRT TELECOM RADIO ELECTR ;PHILIPS ELECTRONICS NV (NL)) 20 December 1995 (1995-12-20) column 1, line 24 - column 2, line 8 -----	1,11,20

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter.   onal Application No

PCT/US 99/05731

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0688121 A	20-12-1995	FR 2721465 A	22-12-1995
		JP 8008975 A	12-01-1996
		US 5649091 A	15-07-1997